Solution Structures of Casein Peptides: NMR, FTIR, CD, and Molecular Modeling Studies of α_{s1}-Casein, 1–23

Edyth L. Malin,[1,4] Michael H. Alaimo,[1] Eleanor M. Brown,[1] James M. Aramini,[2] Markus W. Germann,[2] Harold M. Farrell, Jr.,[1] Paul L. H. McSweeney,[3] and Patrick F. Fox[3]

To determine its potential for interacting with other components of the casein micelle, the N-terminal section of bovine α_{s1}-casein-B, residues 1–23, was investigated with nuclear magnetic resonance (NMR), Fourier transform infrared (FTIR) and circular dichroism (CD) spectroscopies, and molecular modeling. NMR data were not consistent with conventional α-helical or β-sheet structures, but changes in N–H proton chemical shifts suggested thermostable structures. Both CD and FTIR predicted a range of secondary structures for the peptide (30–40% turns, 25–30% extended) that were highly stable from 5°C to 25°C. Other conformational elements, such as loops and polyproline II helix, were indicated by FTIR only. Molecular dynamics simulation of the peptide predicted 32% turns and 27% extended, in agreement with FTIR and CD predictions and consistent with NMR data. This information is interpreted in accord with recent spectroscopic evidence regarding the nature of unordered conformations, leading to a possible role of α_{s1}-casein (1–23) in facilitating casein–casein interactions.

KEY WORDS: α_{s1}-Casein; casein structure; CD; FTIR; milk proteins; molecular modeling; NMR; peptide.

1. INTRODUCTION

Caseins, the principal proteins in milk, have two major functions. Their most commonly understood role is to provide calcium transport for neonatal nutrition. An underlying and less apparent function involves interaction with other caseins to form a framework of submicelles and micelles that will be stable during transit but then readily degraded during digestion (Kumosinski *et al.*, 1994*a*; Schmidt, 1982). For example, strong protein–protein interaction sites occur on the C- and N-terminal sections of α_{s1}-casein-B (Farrell *et al.*, 1988; Horne, 1998; Kumosinski and Farrell, 1991) that would promote formation of a framework structure. Studies on the C-terminal portion of α_{s1}-casein (Alaimo *et al.*, 1999*a*, *b*) have shown a strong propensity for salt-induced self-association. In contrast, correlation of physical studies with molecular modeling (Farrell *et al.*, 2001*a*) has indicated that interactions of the N-terminal peptide (residues 1–23), unlike those at the C-terminal, would be absent at high ionic strength.

Caseins have been described as loose, "flowing" polypeptides with no defined structure (Holt and Sawyer, 1993). However, Raman spectroscopy has demonstrated

Mention of brand or firm name does not imply endorsement by the U.S. Department of Agriculture over others of a similar nature not mentioned.

[1] Eastern Regional Research Center, ARS, USDA, Wyndmoor, Pennsylvania 19038.

[2] Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania 19107.

[3] Department of Food Chemistry, University College, Cork, Ireland.

[4] To whom correspondence should be addressed. e-mail: emalin@arserrc.gov

[5] Abbreviations: CD, Circular dichroism; CSI, chemical shift index; FTIR, Fourier transform infrared spectroscopy; HSQC, heteronuclear single quantum coherence; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser enhancement spectroscopy; PIPES, piperazine-*N*,*N*′-bis(2-ethanesulfonic acid); PPII, poly-L-proline II (all *trans*); RMS, root mean square; SAXS, small-angle X-ray scattering; TOCSY, total correlation spectroscopy.

significant amounts of secondary structure, up to 30% extended (β-sheet-like) conformations and a similar percentage of turns (Byler *et al.*, 1988). Attempts to crystallize caseins for X-ray analysis have not been successful, a formidable obstacle to understanding the factors responsible for the formation of casein micelles and their variable pH- and temperature-dependent behavior in solution (Farrell *et al.*, 2001*a;* Holt, 1992; Schmidt, 1982).

Many instrumental, physicochemical, and theoretical approaches have been used to characterize caseins, casein micelles, and casein peptides including nuclear magnetic resonance (NMR)[5] (Huq *et al.*, 1995; Kakalis *et al.*, 1990; Rollema and Brinkhuis, 1989; Tsuda *et al.*, 1991), circular dichroism (CD)[5] (Chaplin *et al.*, 1988; Creamer *et al.*, 1981), fluorescence (Clarke and Nakai, 1971), infrared (FTIR)[5] and Raman spectroscopy (Byler *et al.*, 1988; Curley *et al.*, 1998), electron microscopy (Kumosinski *et al.*, 1996), and small-angle X-ray scattering (SAXS)[5] (Kumosinski *et al.*, 1994*b*). Several models of the casein micelle have been proposed that combine theoretical and experimental data (Farrell *et al.*, 2001*a;* Holt, 1992; Schmidt, 1982).

Molecular modeling, based on sequence data and aided by solution studies, is recognized as an option for predicting structures of proteins that have neither crystal structure nor sequence homology with other proteins (Pederson and Moult, 1995). This technique was successfully applied to the structures of α_{s1}-, β-, and κ-caseins (Kumosinski *et al.*, 1993*a, b,* 1994*c*), and the predicted structures correlate well with physical properties of associated caseins (Kumosinski *et al.*, 1994*a, b,* 1996). Molecular dynamics simulations of α_{s1}-casein peptides (Malin and Brown, 1995, 1999) showed that their ability to adopt new backbone conformations was variable. Most acquired more secondary structure, but the few that resisted conformational changes are located at the N- and C-terminal regions of α_{s1}-casein and are likely to be responsible for its open, flexible structure.

The structure of a specific α_{s1}-casein peptide, residues 1–23, which is formed by the action of chymosin, is of particular interest. The N-terminal portion of the α_{s1}-molecule has been proposed as a reactive site in self-association reactions of the molecule (Farrell *et al.*, 2001*a;* Horne, 1998), as noted above. In addition, the region of α_{s1}-casein (1–23) surrounding residue 20 has been suggested to contain a significant allergenic epitope recognized by human IgE (Spuergin *et al.*, 1996). Finally, the fragment composed of the first nine residues of α_{s1}-casein is resistant to further cleavage by enzymes of starter culture bacteria; it accumulates during cheese ripening (Fox *et al.*, 1995) and contributes to bitter flavor in cheese (Pederson *et al.*, 1999). Preliminary CD analysis of α_{s1}-casein (1–23) was reported by Chaplin *et al.*, (1988).

In the research described here, four different techniques were used to define a solution structure for the peptide: NMR, CD, FTIR, and molecular modeling.

2. METHODS

2.1. Materials

All reagents used were of analytical grade or ACS certified from Sigma (St. Louis, MO).

2.2. Peptide Synthesis

The first 23 residues of α_{s1}-casein were synthesized (National Food Biotechnology Centre, Cork, Ireland) on a 0.25-mmole scale using a Model 431A peptide synthesizer and FastMoc synthesis with HMP resin (Applied Biosystems, San Jose, CA). After the peptide was cleaved from the resin, it was precipitated with ether and lyophilized to remove traces of synthesis and cleavage reagents. Mass spectrometric analysis after synthesis confirmed that the peptide had the correct sequence. Amino acid analysis (Waters Pico-Tag, Milford, MA) showed the composition to be identical to that of native peptide, and sequence analysis (Model 437 Sequencer, Applied Biosystems, San Jose, CA) indicated the correct order of amino acids.

^{19}F NMR spectroscopy showed that approximately 4–5 moles of trifluoroacetic acid (TFA) per mole of peptide remained bound to the peptide after synthesis. Removal of TFA was accomplished by treating a solution of the peptide in 10 mM PIPES buffer with 0.2 M trimethylamine to a pH of 8.5. After lyophilization, the peptide was dissolved in 50 mM NH_4HCO_3 and chromatographed on a Sephadex G-50 column. The peptide was recovered and lyophilized; when retested by ^{19}F NMR, less than 0.5 mole TFA/mole peptide remained.

2.3. Analytical Ultracentrifugation

For analytical ultracentrifugation, the peptide samples were dissolved at concentrations ranging from 1.0 to 3.0 mg/ml. Samples and solvents were filtered with a 0.45-μm membrane filter; less than 1% of the material was retained on the filter as ascertained by UV spectroscopy. Sedimentation equilibrium experiments were performed using a Beckman Optima XL-A (Palo Alto, CA) analytical ultracentrifuge at speeds ranging from 26,000 to 46,000 rpm at 25°C. A 12-mm charcoal-epon

6 channel centerpiece was used with quartz windows in wide-aperture window holders.

The solvent densities used in these experiments were 1.0016 (μ = 0.020 M) and 1.0060 (μ = 0.100 M) for the low- and high-ionic-strength experiments, respectively. The partial specific volume of the peptide was 0.73 ml/g, calculated from amino acid composition. The calculated molar absorptivity was 1501 $mole^{-1}$ cm^{-1} at 260 nm. Data were collected at 260 nm using the standard XL-A procedure. The absorption versus radius plots were analyzed directly for weight-average molecular weight using the program IDEAL 1, a part of the Optima XL-A data analysis software. As the absorbance offsets were not allowed to float in these analyses, weight-average molecular weights were obtained. Analysis of the data was accomplished using ASSOC4, which is designed for a system of up to four species. This model represents a parallel association scheme where monomer is simultaneously in equilibrium with dimer, trimer, and tetramer when n_2, n_3, and n_4 are the integers 2, 3, and 4 in Eq. (1) below. For α_{s1}-casein (1–23) after removal of TFA, the best fits were obtained by fixing the sequence molecular weight at 2770 for the monomer and floating K values at the increasing integer values of n noted above. Analysis of the data was accomplished using the following equation:

$$\begin{aligned} A_T = {} & \exp[\ln(A_0)] + HM(X^2 - X_0^2)] \\ & + \exp[n_2 \ln(A_0) + \ln(K_{a2}) + n_2 HM(X^2 - X_0^2)] \\ & + \exp[n_3 \ln(A_0) + \ln(K_{a3}) + n_3 HM(X^2 - X_0^2)] \\ & + \exp[n_4 \ln(A_0) + \ln(K_{a4}) + n_4 HM(X^2 - X_0^2)] \\ & + E \end{aligned} \quad (1)$$

where A_T is the total absorbance of all species at radius X, A_0 is the absorbance of the monomer species at reference radius X_0, H = constant · $[(1 - \nabla\rho)\omega^2]/2RT$, M is the apparent monomer molecular weight, n_i is the stoichiometry for species i (number of monomers), K_{ai} is the association constant for the monomer–$n$mer equilibrium of species i, and E is the baseline offset.

As determined in Eq. (1), K_{ai} may be converted to molar units using

$$K_{\text{conc}} = K_{\text{abs}}\left(\frac{\varepsilon l}{n}\right)^{n-1} \quad (2)$$

where K_{abs} is K_{ai} of the software, ε is the molar absorptivity, l is the path length in centimeters, and n is the fitted integer of the software. In turn the molar constant K_{conc} can be converted to the weight constant k_n by

$$k_n = \frac{K_{\text{conc}}}{(M_{\text{monomer}}/n)^{n-1}} \quad (3)$$

Here M_{monomer} is the monomer molecular weight and n has the same meaning as above. The k_n values reported here were determined in this fashion.

2.4. NMR Spectroscopy

Samples for NMR spectroscopy were prepared in 0.4 ml of 10 mM sodium phosphate–70 mM NaCl, pH 5.0, 90% H_2O/10% D_2O. Peptide concentrations were 2.0 and 13.4 mM; final concentrations were determined using the calculated molar absorptivity of 1501 $mole^{-1}$ cm^{-1} at 260 nm. An additional 20 μM peptide sample was prepared in 99.9% D_2O for a dilution study. The pH of the 2.0 mM sample was adjusted between 2.8 and 7.0 by adding NaOD and/or DCl. Spectra for the 2.0 mM, pH 5.0, sample were obtained over a temperature range of 10–45°C in order to monitor the temperature dependence of amide proton chemical shifts.

All NMR experiments were performed on a Bruker AMX 600 NMR spectrometer equipped with a 5-mm broadband inverse probe, at ^1H and ^{13}C frequencies of 600.1 and 150.9 MHz, respectively, using the XWIN-NMR 1.1.1 software package. Typically, 1D ^1H spectra were obtained with a total repetition time of 2.5–3.5 sec and presaturation of the H_2O solvent signal. All 1D ^1H NMR data were Fourier transformed with 0.3- to 1.0-Hz exponential line broadening. Phase-sensitive 2D nuclear Overhauser enhancement spectroscopy (NOESY) and total correlation spectroscopy (TOCSY) experiments were obtained using TPPI (Marion and Wüthrich, 1983). Mixing times were 300 and 124 ms for the NOESY and TOCSY experiments, respectively. Heteronuclear shift correlations were obtained using the ^1H and ^{13}C heteronuclear single quantum coherence (HSQC) experiment (Wider and Wüthrich, 1993). Data from 2D experiments were processed with 90°-shifted (for TOCSY and HSQC) and 45°-shifted (for NOESY) sine functions in both dimensions, followed by automatic baseline correction in F1 and F2. The ^1H and ^{13}C chemical shift values were referenced to external 2,2-dimethyl-2-silapentane-5-sulfonate (DSS).

2.5. Infrared (FTIR) Spectroscopy

For infrared (FTIR) measurements, 3–5 mg of TFA-free α_{s1}-casein (1–23) was dissolved as a 1.5% w/w solution in 10 mM sodium phosphate buffer–70 mM NaCl, pH 5.0 (μ = 0.100). Spectra were obtained at 25°C using a Nicolet 740 FTIR spectrometer (Madison, WI) equipped with a Nicolet 660 data system. Data collection was

preceded by a nitrogen purge of the sample chamber, which consisted of a demountable cell with CaF_2 windows separated by a 6-μm Teflon spacer. Each spectrum consisted of 4096 double-sided interferograms, coadded, phase-corrected, apodized with a Happ–Genzel function, and fast-Fourier-transformed. Secondary structural features were calculated from the amide I envelope, after assignment of peaks by use of second derivative spectra and fitting of Gaussian peaks to the original spectra. The selected peaks were refitted to the original spectra using an iterative curve-fitting procedure, as previously reported (Kumosinski and Unruh, 1996). Data are reported as the sum of assigned spectral elements gathered from the 20-protein database of Kumosinski and Unruh (1996).

2.6. Circular Dichroism (CD) Spectroscopy

CD experiments were carried out with 0.3-mg/ml peptide samples of TFA-free α_{s1}-casein (1–23) in a 10 mM sodium phosphate–70 mM NaCl buffer at pH 5.0, $\mu = 0.100$. Samples were filtered through 0.45-μm-pore filters into a jacketed 0.5-mm-pathlength cylindrical cell made of far-UV quartz. CD spectra were recorded on an AVIV 60DS spectropolarimeter (Aviv Associates, Lakewood, NJ) calibrated with *d*-10-camphorsulfonic acid. Successive measurements in the far UV (250–195 nm) were made at 25°C, 10°C, and 5°C. Ellipticities are reported as deg cm^2 $dmol^{-1}$. Secondary structure was estimated using the linear combinations described by Greenfield (1996) with the databases of Yang *et al.* (1986) and Brahms and Brahms (1980). Data were also analyzed using the CONTIN method of Provencher and Glöckner (1981).

2.7. Molecular Modeling

The structure of α_{s1}-casein (1–23) was simulated with the molecular modeling program SYBYL, Version 6.0 (Tripos Associates, St. Louis, MO), starting with the refined model of α_{s1}-casein (Kumosinski *et al.*, 1994*c*). Selection of appropriate conformational states for individual amino acid residues in that model was previously described (Kumosinski *et al.*, 1994*c*). After using SYBYL subroutines to excise residues 24–199 from the parent chain, the remaining peptide (residues 1–23) retained the identical conformation of the sequence in the intact refined protein. An N-methyl blocking group was added to the carboxyl terminus, and a droplet of 600 molecules of water was introduced surrounding the peptide to simulate solution conditions. Energy minimizations and dynamics were based on Kollman charges (Weiner *et al.*, 1984, 1986) and the Tripos force field (Clark *et al.*, 1989). Molecular dynamics simulations were conducted over a range of temperatures from 50 to 300 K in 50 K increments, each followed by energy minimization. Several analytical subroutines of SYBYL were used to compare changes in structure that occurred during these operations. These included secondary structure formed or lost and distance between the α-carbons of the amino and carboxyl termini.

3. RESULTS AND DISCUSSION

3.1. Analytical Ultracentrifugation Studies of α_{s1}-Casein (1–23)

The initial studies of the α_{s1}-casein (1–23) peptide were conducted at low salt concentration (10 mM sodium phosphate, pH 5.0), prior to the removal of bound TFA as described above. The weight-average molecular weight was found to be 5110, indicating a preponderance of associated species (sequence weight = 2770). However, when the data were analyzed by Eq. (1), the monomeric species was calculated to be 1460 ± 90 (Table I). Several possibilities could explain the result. Proteolysis was ruled out, as sequence analysis of a solution of the peptide under comparable conditions gave a linear repetitive yield ranging from 260 to 150 pmoles and no new N-terminals. From the physical chemical perspective, the lack of supporting electrolyte (a low salt concentration) may reduce the observed molecular weight by a factor of $1/(n + 1)$, where n is the potential charge on the molecule; another possibility is bound salt (Schachman, 1959).

Ultracentrifugation, following removal of TFA, gave a weight-average molecular weight of 6180. However, a value of 2770 for the monomeric species was derived from Eq. (1). The calculated value of n (the degree of self-association) derived from Eqs. (1)–(3) also changed (Table I) and k_2 increased significantly. All of the above data point to a high degree of self-association at low salt concentration (dimer and above). However, because of the possibility of bound salt and the low concentrations of supporting electrolyte (Schachman, 1959), the values for k_2 and n must be viewed with extreme caution.

When the peptide was subjected to analytical ultracentrifugation in 10 mM phosphate–70 mM NaCl, pH 5.0, $\mu = 0.100$, the data could be readily fit to an ideal solute model with a molecular weight equal to the sequence weight of 2770 ± 6%. At this ionic strength, both TFA-containing samples and treated samples gave similar results (Table I).

Table I. Effect of Increased Salt on the Association Constants and Apparent Molecular Weights of α_{s1}-Casein (1–23) by Ultracentrifuge Analysis

Sample[a]	Sedimentation equilibrium		
	MW_w[b] (Da)	$k_{a2} + n_2$ (ml/g)	Calc. monomer weight(Da)
Low salt: μ = 0.020			
α_{s1}-Casein (1–23) untreated	5110 ± 409	0.45 ± 0.03 (10)[c,d]	1460 ± 90
α_{s1}-Casein (1–23) treated	6180 ± 618	6.39 ± 0.32 (4)[c,e]	2770 ± 70
High salt: μ = 0.100			
α_{s1}-Casein (1–23) treated	2770[f]		

[a] Untreated, as synthesized; treated, 4–5 moles of TFA removed.
[b] Weight-average molecular weight; four determinations at 20°C.
[c] Conversion to k_2 as described in Section 2.
[d] Average of four experiments, each containing two samples; $n = \pm 1$.
[e] Average of three samples in one experiment; $n = \pm 1$.
[f] Three determinations of two samples (six sets of data) at 25°C; one sample not chromatographed after trimethylamine treatment gave a similar value.

Based on molecular modeling studies of α_{s1}-casein (Kumonsinski *et al.*, 1994*c*), electrostatic bonds could be formed between Arg22 and Glu14 on adjacent monomers by docking their N-terminal regions in an antiparallel fashion. These intermolecular bonds are in a relatively hydrophobic environment and thus would have somewhat added bond strength (Tanford, 1967). Farrell *et al.* (1988) demonstrated that α_{s1}-B, the predominant genetic variant, is not readily salted in, whereas α_{s1}-A, which lacks residues 14–26, is soluble at 1°C. A fivefold difference in salting-in constant (A > B) was calculated for the two genetic variants (Farrell *et al.*, 1988). If such intermolecular bonds are formed when two α_{s1}-B chains are docked, they may also occur when two or more chymosin-generated peptides, α_{s1}-casein (1–23), associate. Increased ionic strength would lead to dissociation of the intermolecular salt bridges involving Glu14 and Arg22 residues predicted for docked models of α_{s1}-B, and this was observed. Thus, the salt-induced self-association occurs at low ionic strength but is abolished at elevated ionic strength.

3.2. NMR-Studies of α_{s1}-Casein (1–23)

Measurements of the 1D proton spectrum for α_{s1}-casein (1–23) were made for the pH range 2.8–7.0 at temperatures between 10°C and 45°C to determine conditions for optimum resolution of amide proton signals. Resolution of backbone amide proton signals was best accomplished at pH 5.0 (Fig. 1).

The α_{s1}-casein (1–23) peptide exists predominantly as a monomer in the presence of 70 mM NaCl, as observed by 1D ^{1}H NMR for a dilution series of the peptide. The number of peaks, their chemical shifts and linewidths, with the exception of exchangeable protons which are not observed in D_2O, are nearly identical for solutions of 13.4 and 2.0 mM peptide in H_2O and 20.0 μM peptide in D_2O. These observations agree with results from analytical ultracentrifugation, noted above.

The assignment of the ^{1}H spectrum was carried out for the peptide at 10°C, pH 5.0, using TOCSY, ^{1}H–^{13}C HSQC, and NOESY 2D NMR experiments. The chemical shift values measured for protons and carbons of α_{s1}-casein (1–23) are given in Table II. The NH–H_α

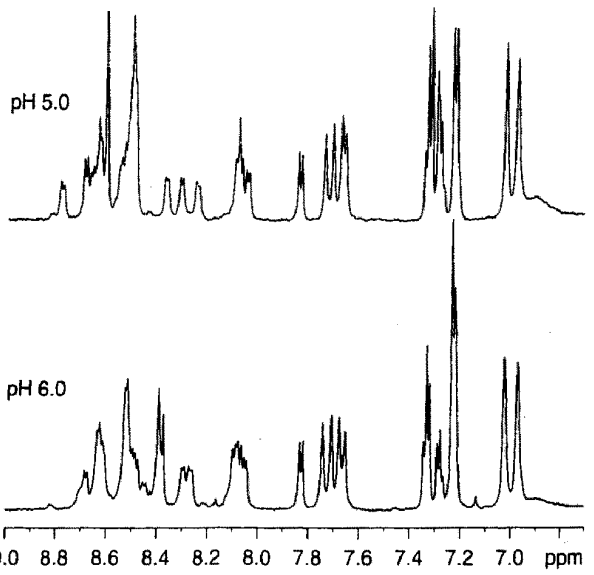

Fig. 1. Amide proton region of the ^{1}H (600.1 MHz) NMR spectrum of 2.0 mM α_{s1}-casein (1–23) in 10 mM Na_2HPO_4–70 mM NaCl, 90% H_2O/10% D_2O. Top: pH 5.0, 10°C; Bottom: pH 6.0, 10°C.

Table II. ^1H and ^{13}C Resonance Assignments for α_{s1}-Casein (1–23)[a]

Residue	Chemical shift (ppm) NH	$C_\alpha H$	C_α	Other
Arg1	8.069[b]	4.282	54.08	$C_\delta H_2$ 3.145, $N_\epsilon H$ 7.216
Pro2	—	4.486	63.09	—
Lys3	8.684	4.174	—	$C_\epsilon H_2$ 2.356
His4	8.500	4.053	—	—
Pro5	—	4.452	63.13	—
Ile6	8.483	4.134	61.08	—
Lys7	8.519	4.285	—	$C_\epsilon H_2$ 2.977
His8	8.780	4.469	53.48	—
Gln9	8.662	4.333	—	$C_\beta H_2$ 1.965, 2.066, $C_\gamma H_2$ 2.341, 2.343
Gly10	8.635	3.946	44.83	—
Leu11	8.299	4.629	55.83	—
Pro12	—	4.394	63.57	—
Gln13	8.624	4.216	—	$C_\gamma H_2$ 2.368, 2.370
Glu14	8.502	4.255	56.83	$C_\beta H_2$ 1.987, $C_\gamma H_2$ 2.267
Val15	8.232	4.042	62.65	$C_\beta H$ 2.044, $C_\gamma H_3$ 0.911
Leu16	8.360	4.311	—	—
Asn17	8.495	4.637	—	$C_\beta H_2$ 2.765, 2.845
Glu18	8.547	4.195	—	—
Asn19	8.493	4.640	53.29	$C_\beta H_2$ 3.221, 3.099
Leu20	8.042	4.284	—	—
Leu21	8.081	4.300	—	—
Arg22	8.065	4.384	—	$C_\delta H_2$ 3.235, $N_\epsilon H$ 7.284
Phe23	7.831	4.483	58.98	$C_\beta H_2$ 3.154, 2.926

[a] Missing carbon shifts could not be unambiguously assigned.
[b] Terminal NH_3^+.

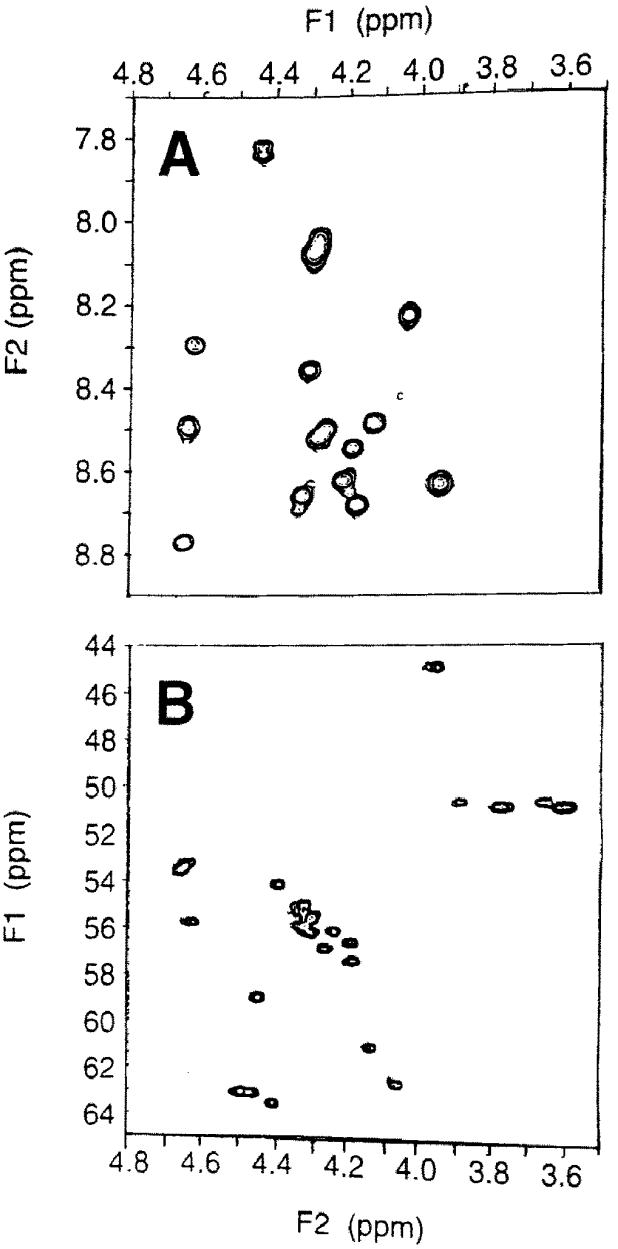

Fig. 2. The H_N–H_α fingerprint region of the 600.1-MHz TOCSY (**A**) and the C_α–H_α region of the ^1H–^{13}C HSQC (**B**) spectrum of 2.0 mM α_{s1}-casein (1–23) in 10 mM Na_2HPO_4–70 mM NaCl, 90% H_2O/10% D_2O, pH 5.0, 10°C.

fingerprint region of the TOCSY spectrum is shown in Fig. 2A and the C_α–H_α region for the HSQC spectrum in Fig. 2B. In the TOCSY spectrum the expected cross-peaks were observed for a peptide of 23 residues containing three prolines. In addition, in spite of the overlap, the appropriate number of C_α–H_α cross-peaks were observed in the HSQC spectrum. The three sets of peaks at 50.57–50.80 ppm in the carbon (F1) dimension correspond to the proline C_δ–H_δ correlation. These were assigned to the *trans* isomers of Pro2, Pro5, and Pro12, based on reported proline chemical shift values for a large series of short, proline-containing linear peptides (Dyson *et al.*, 1988). Crosspeaks due to *cis* forms were not readily apparent, in agreement with the data of Kakalis *et al.* (1990), who found >85% of proline residues in whole casein to be in the *trans* form by ^{13}C NMR. Taken together, these observations indicate the absence of slowly interconverting conformers.

A number of unique spin systems were readily identified from characteristic patterns in the full TOCSY spectrum, including Gly10, Leu11, Val15, and Leu16. Sequential assignments between these starting points were obtained using the d_{NN} (i, i + 1) and $d_{\alpha N}$(i, i + 1) cross-peaks in the NOESY spectrum. An expansion of the fingerprint region showing all sequential $d_{\alpha N}$(i, i + 1) contacts is given in Fig. 3. The entire sequence of residues can be traced through the short-range $d_{\alpha N}$(i, i + 1) NOE network, with the exception of the N-terminal residue (Arg1) and the three proline residues which interrupt the sequential backbone assignments. An expansion of the amide region of the NOESY spectrum showing d_{NN}(i, i + 1) cross-peaks is given in Fig. 4, where sequential connectivities for a number residues are shown.

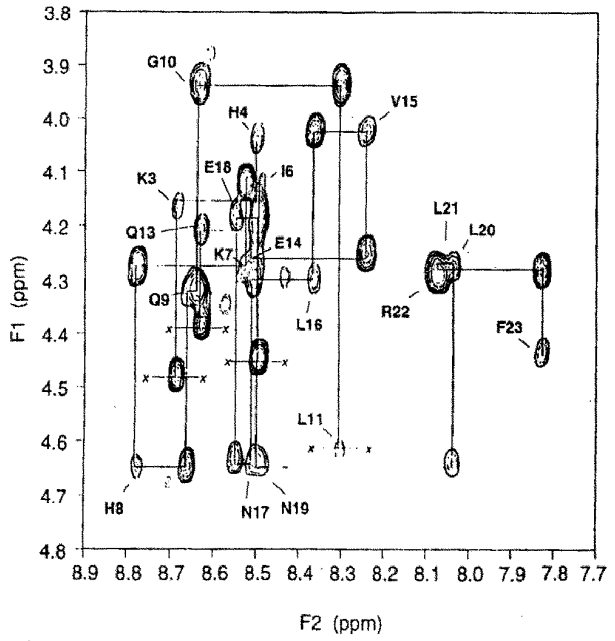

Fig. 3. The H_N–H_α fingerprint region of the 600.1-MHz NOESY spectrum (mixing time = 300 msec) of 2.0 mM α_{s1}-casein (1–23) in 10 mM Na_2HPO_4–70 mM NaCl, 90% H_2O/10% D_2O, pH 5.0, 10°C. Sequential connectivities are labeled for each residue. The horizontal lines marked on either side with × represent H_N–H_α sequential connectivities broken by proline residues.

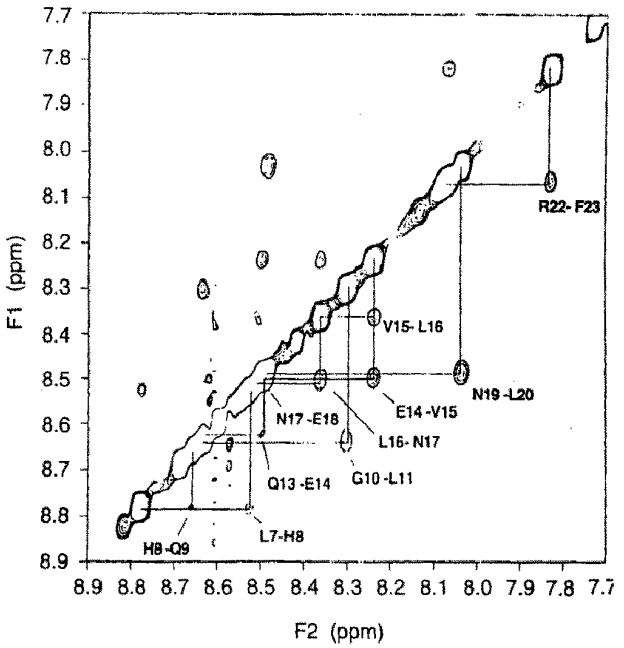

Fig. 4. The backbone amide proton region of the 600.1-MHz NOESY spectrum (mixing time = 300 msec) of 2.0 mM α_{s1}-casein (1–23) in 10 mM Na_2HPO_4–70 mM NaCl, 90% H_2O/10% D_2O, pH 5.0, 10°C. Sequential amide–amide connectivities are labeled for each resolved pair of residues.

NMR data are summarized in Fig. 5. The solution conformation of the peptide was explored by determination of medium/long-range NOE couplings, three-bond spin–spin (J) coupling constants, C_α proton chemical shift values, and the temperature dependence of the amide proton chemical shift. The most useful data for determining peptide folding patterns are NOE values, which depend upon the interproton distance r_{ij}[6]. Medium-range $d_{\alpha N}$ (i, i + 3) and $d_{\alpha\beta}$ (i, i + 3) NOEs are often diagnostic of helices. The presence of a four-residue, hydrogen-bonded reverse turn (Dyson *et al.*, 1988) is denoted by strong d_{NN} between residues 3 and 4 as well as strong $d_{\alpha N}$ between residues 2 and 4. No NOE couplings beyond the sequential (i, i + 1) were observed for the α_{s1}-casein (1–23) peptide (Fig. 5).

Three-bond $J_{\alpha N}$ coupling constants for extended chains are 6–9 Hz. Smaller coupling constants (2–5 Hz) are observed for α-helices and for the second and/or third residues of β-turns (Dyson *et al.*, 1988). All $^3J_{\alpha N}$ constants measured for α_{s1}-casein (1–23) were in the 6.6- to 8.8-Hz range, with the exception of 5.3 Hz for Leu16 (Fig. 5).

The chemical shift index (CSI), developed by Wishart and Sykes (1994), compares each $C_{\alpha H}$ chemical shift in a peptide/protein to a range of random coil values, as determined from a large database of peptides and proteins (Wishart *et al.*, 1991). A chemical shift value lower than the range may indicate that the residue lies in a region of α-helix, while a value greater than the range may indicate participation in a β-strand extended structure.

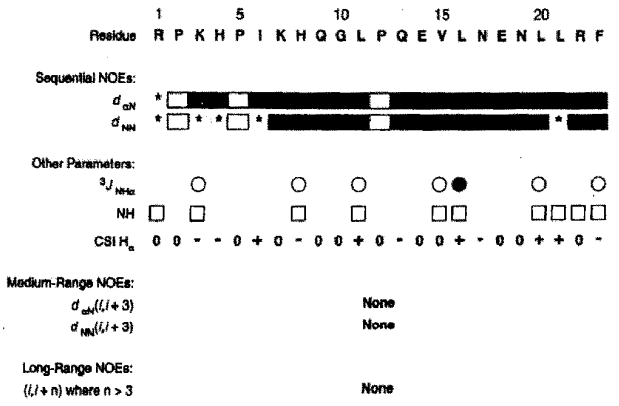

Fig. 5. Summary of NMR spectral data for α_{s1}-casein (1–23). The gray horizontal bars represent sequential $d_{\alpha N}(i, i + 1)$ and $d_{NN}(i, i + 1)$ NOE couplings. Cross-hatched bars represent sequential connectivities broken by proline residues. Asterisks denote NOE cross-peaks that were not resolved due to spectral overlap. A $^3J_{\alpha N}$ coupling >6.5 Hz is indicated by an open circle, while the black circle represents a $^3J_{\alpha N}$ coupling <5.5 Hz. Open squares indicate an amide temperature coefficient ($-\Delta\delta/\Delta T$) in the range 8–11 ppb/K. Chemical shift indices (CSI) are derived for low-ppm (upfield) shifts (−) or high-ppm (downfield) shifts (+).

A grouping of four or more upfield shifts is designated as an α-helix, while three or more consecutive downfield shifts imply β-strand (Wishart *et al.*, 1992). While several $C_{\alpha H}$ chemical shifts fall outside the random coil range (Fig. 5), the data do not suggest local regions of classical α-helix or β-structure.

The temperature dependence of amide proton chemical shifts may be strongly affected by hydrogen bonding. Low-temperature coefficients (<4.0 ppb/K) can result from interresidue hydrogen bonds in α-helices and reverse turns and interstrand hydrogen bonding in β-pleated sheets (Dyson *et al.*, 1988). All amide temperature coefficients for α_{s1}-casein (1–23) in Fig. 5 were in the 8–11 ppb/K range, indicating that those residues are solvent exposed and not involved in conventional α-helix or extended β-sheet. However, when the chemical shifts of the amide protons are viewed relative to those found for random coil, a pattern emerges for this peptide (Wishart and Sykes, 1994). Table III shows calculated Δ(N–H) values and their ratios to the average chemical shift for amide protons to either β-sheet or α-helix. The Δ(N–H) values for residues 3, 4, and 6–10 are indicative of β-sheet, but the $C_{\alpha H}$ indices in Table II are not. Although not hydrogen bonded, this area may represent extended β-sheet. A plot of Δ(N–H) values in Fig. 6 suggests a sheet including residues 3–15, interrupted by Pro5 and Pro12. Note that residues 20–23 have negative Δ(N–H) values, which suggests the potential for nascent helix.

Taken together, the NMR evidence from medium-range NOEs, $^3J_{\alpha N}$, CSI, and temperature coefficients support the conclusion that the peptide in solution does not adopt conventional helical or sheet structures. This indicates that on the NMR time scale the peptide has a narrow range of rapidly interconverting conformational states. The relatively large number of proline residues (12%) in this small peptide may diminish its ability to form α-helices and β-sheet near the N-terminal end, yielding a polyproline-like structure instead. Proline residues may have some tendency to promote turns in α_{s1}-casein (1–23), but these do not display typical hydrogen bonding patterns. The *trans* nature of the proline residues may impart a stiffness to the peptide, preventing large conformational fluctuations. Ramachandran calculations have shown that an individual amino acid residue in a random coil peptide spends up to 95% of its time in the broad β-sheet region (Feeney *et al.*, 1972). These β-extended structures (turns, sheets, etc.) that are observed in the β region of ϕ, ψ space, may be represented on the NMR time scale by structures with N–H chemical shifts such as those shown in Fig. 6. Such structures could occupy the upper left quadrant of Ramachandran plots and may be better identified by CD and FTIR.

Table III. Change in Chemical Shifts for ^{1}H–N NMR for α_{s1}-Casein (1–23) Relative to Random Coil Peptides

Residue	Δ N-H for α_{s1}(1–23)	Δ(N–H) Ratio[b]
Arg1	—[c]	—[c]
Pro2	—	—
Lys3	0.43	1.5
His4	0.22	0.8
Pro5	—	—
Ile6	0.27	0.9
Lys7	0.27	0.9
His8	0.50	1.7
Gln9	0.39	1.3
Gly10	0.34	1.2
Leu11	0.06	0.2
Pro12	—	—
Gln13	0.35	1.2
Glu14	0.14	0.5
Val15	0.04	0.1
Leu16	0.13	0.5
Asn17	0.09	0.3
Glu18	0.19	0.6
Asn19	0.11	0.4
Leu20	−0.19	1.0
Leu21	−0.15	0.8
Arg22	−0.21	1.1
Phe23	−0.45	2.3

[a] $\Delta = ppm_{obs} - ppm_{rc}$; where ppm_{obs} is from Table II and ppm_{rc} is from Wishart and Sykes (1994), Table VII.

[b] Ratio of Δ for the peptide to the average positive change in chemical shift (+ 0.29 for β-structures) or negative chemical shift (− 0.19 for α-helices) (Wishart and Sykes, 1994, Table VII).

[c] No chemical shift for an amino terminal.

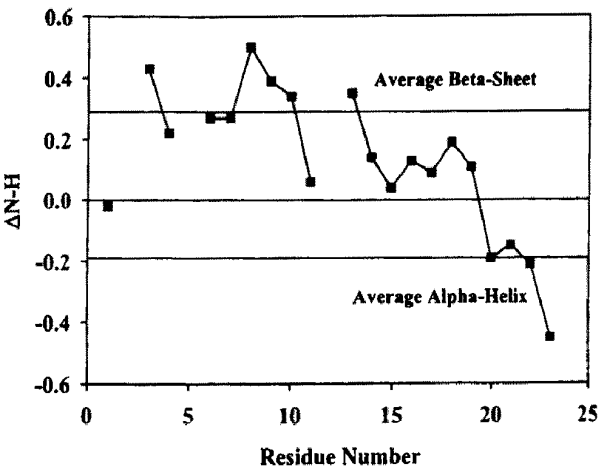

Fig. 6. Plot of Δ(N–H) ($ppm_{obs} - ppm_{rc}$) for α_{s1}-casein (1–23), as defined in Table III. The upper line represents the average positive shift for β-sheet, the lower line represents the average negative shift for α-helix.

3.3. Estimation of Secondary Structure of α_{s1}-Casein (1–23) by FTIR

Global secondary structure was estimated from the FTIR spectra of a 1.5% w/w solution of TFA-free α_{s1}-casein (1–23) at pH 5.0 and 5°C in 10 mM sodium phosphate–70 mM NaCl (μ = 0.100), in which the peptide is a monomer (Fig. 7A). Figure 7B shows the fit of multiple Gaussian peaks to the original spectrum using procedures previously developed (Curley *et al.*, 1998; Kumosinski and Unruh, 1996).

Tentative peak assignments for these bands were made using the results of Krimm and Bandekar (1986) as a primary guide, and refinements were based upon studies with globular proteins (Kumosinski and Unruh, 1996).

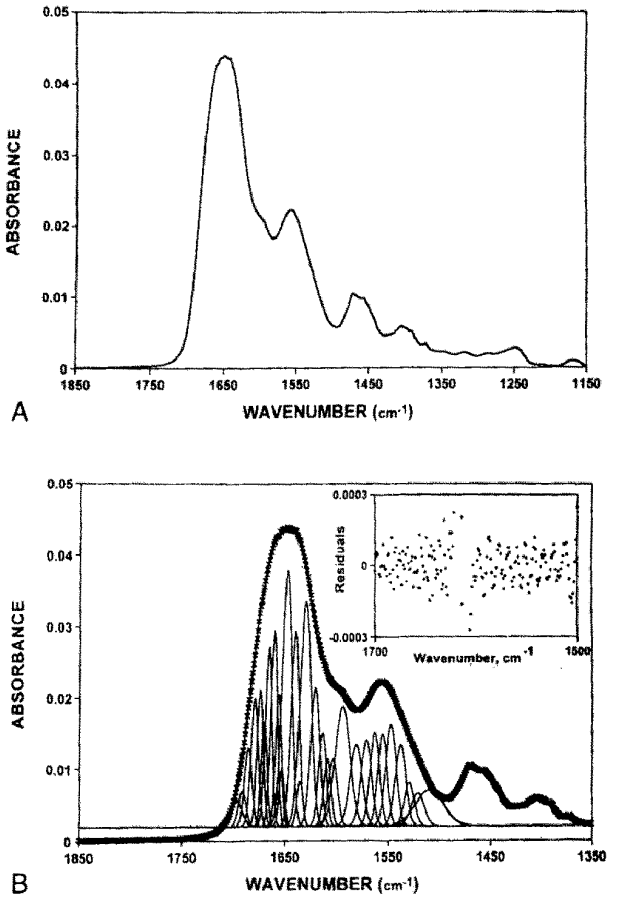

Fig. 7. (A) FTIR spectrum of α_{s1}-casein (1–23) in 10 mM sodium phosphate–70 mM NaCl, pH 5.00, 5°C. Frequency in wavenumbers. (B) Using second derivative analysis of the FTIR spectrum of α_{s1}-casein (1–23) at 5°C, amide I and II regions were fitted at peak positions with Gaussian curves. Following the initial assignments, nonlinear regression analysis was used to fit the data. Crosses represent experimental data; solid lines represent fitted individual Gaussian components and their sum. Inset is the corresponding residual plot obtained by comparison with the original spectrum.

The amide I and II regions were fitted at peak positions obtained from second derivative analysis with Gaussian curves. The initial assignments for the amide I regions are given in Table IV. Figure 7B shows bands centered close to 1656 cm^{-1}; such bands might be attributed theoretically to elements of α-helix or β-turns (Kumosinski and Unruh, 1996). However, the NMR results rule out classical H-bonded helices and turns. In a study that compared CD, FTIR, and X-ray diffraction data on murine interleukins, Wilder *et al.* (1992) concluded that in the absence of CD evidence for α-helices, FTIR bands centered on 1656 cm^{-1} may be assigned to large loop or extended turn structures. Accordingly, we assigned a separate category termed large loop and helix for the elements centered on 1656 cm^{-1}.

The absence of NMR-detectable α-helix provides additional rationale for assigning the FTIR bands between 1658 and 1648 cm^{-1} (representing 25% of the structure) to loop, turn, or other extended structures. Analysis of globular proteins with known crystal structures has shown that short runs of poly-L-proline II-like

Table IV. Amide I Peak Areas for the FTIR Spectra of α_{s1}-Casein (1–23) at 5°C

Conformational element	Frequency[a] (cm^{-1})	Area (%)	
		$\bar{X}^2$	SE[a]
Turn	1697	1.9	0.1
Turn	1691	1.8	0.5
Turn	1686	3.9	1.0
Turn	1680	6.2	0.5
Turn	1675	5.7	0.8
Turn[b]	1671	3.9	0.5
Turn[b]	1667	7.8	1.0
Turn[b]	1661	8.0	0.8
Large loop[c]	1657	5.2	1.0
α-Helix	1654	2.1	0.3
Loop or helix[d]	1649	16.2	1.2
Irregular	1641	10.7	0.8
β-Sheet[e]	1636	1.8	1.3
β-Sheet[e]	1631	16.3	0.8
β-Sheet[e]	1621	8.5	0.6
RMS[f]	3.4×10^4		

[a] Frequencies ± 2 cm^{-1}. Standard error calculated as in Kumosinski *et al.*, (1997) for reanalysis of the same sample.
[b] Contains elements of amide side chains (Asn and Gln) as well as bent strand and 3–10 helix (Kumosinski and Unruh, 1996).
[c] Designated as arising from large loops (Wilder *et al.*, 1992), but may contain α-helical and β-turn elements (Kumosinski and Unruh, 1996) or polyamino acid-like elements.
[d] Unresolved loop and α-helix in the fitting routine.
[e] These elements include β-sheet as well as extended β-structure (Kumosinski and Unruh, 1996) or PPII (Farrell *et al.*, 2001*b*).
[f] Root mean square of the fit.

Table V. Comparison of Secondary Structure Estimates for α_{s1}-Casein (1–23)

Method	β-Sheet (%)	Turns (%)	Unspecified (%)	α-Helix-loop[a,b] (%)
FTIR[c]	27 ± 3	39 ± 1	11 ± 1	24 ± 1
CD[d]	29 ± 1	31 ± 1	26 ± 1	13 ± 1
Inferred from NMR[e]	27	0	55	18
Molecular modeling[f]	27	32	41	0

[a] For FTIR, includes three bands: 1650 ± 2, 1655 ± 2, and 1658 ± 2 cm^{-1}; the first corresponds to α-helix.
[b] For CD, includes α-helix only.
[c] Average of three determinations in 10 mM sodium phosphate–70 mM NaCl, pH 5.0.
[d] Average fits, one determination (six accumulations at each temperature).
[e] Interpretation of Wishart and Sykes (1994). Using Δ(N–H) ppm and the N–H ratio would yield the upper limit for possible structure from NMR.
[f] Secondary structure predicted by SYBYL analysis of the peptide conformation after molecular dynamics.

structure (PPII) (4–6 residues) do occur quite frequently (Adzhubei and Sternberg, 1993). Dukor and Keiderling (1991) reported that short runs of poly-L-proline (3–6 residues) would yield FTIR spectra with maxima at 1656 and 1620 cm^{-1} in D_2O. Farrell *et al.* (2001*b*) found that in H_2O, PPII (n = 3,4) gave FTIR bands at 1620, 1636, and 1650 cm^{-1} in a ratio of 3:1:1. Table IV shows these bands, suggesting the occurrence of PPII elements in the peptide. These elements are usually not hydrogen bonded and thus not detected by NMR (Adzhubei and Sternberg, 1993). Smyth *et al.* (2001) suggested that PPII-like structures exist in α_s-casein on the basis of a Raman optical activity spectrum.

Values of estimated bulk secondary structure for α_{s1}-casein (1–23) at 5°C were calculated by summing the areas of the FTIR peaks (Table V). The results were as follows: 27% β-extended + PPII structures, 39% turns, 11% unordered, and 23% large loop + helix.

3.4. Circular Dichroism of α_{s1}-Casein (1–23)

For CD measurements, α_{s1}-casein (1–23) was dissolved in 10mM sodium phosphate–70 mM NaCl (μ = 0.100) at pH 5.0, where the peptide is monomeric. Baseline-corrected spectra of α_{s1}-casein (1–23) at 5°C, 10°C, and 25°C are shown in Fig. 8. The absence of minima at 208–212 and 218–222 nm suggests that little α-helix is present, while the weak, broad shoulder at 224–230 nm suggests the presence of β-sheet or PPII, as proposed by Makarov *et al.* (1984). Secondary structure was estimated by two methods. The first was constrained least squares analysis (LINCOMB) (Greenfield, 1996), using the Yang *et al.* (1986) and Brahms and Brahms (1980) databases. Databases for LINCOMB analyses contain reference curves for each type of structure that were extracted from a collection of protein (Yang *et al.*, 1986) or peptide (Brahms and Brahms, 1980) spectra. The second method, CONTIN analysis (Provencher and Glöckner, 1981), has a protein database that also includes polyglutamic acid (Manavalan and Johnson, 1987). The data at 5°C are compared with FTIR data and with the NMR upper structural limits in Table V.

A single three-residue turn of α-helix is suggested by the CONTIN method, in which the CD spectrum is fit directly, using a linear combination method and a large database of proteins of known conformations. The reference curves for α-helix used in the LINCOMB analyses are helix length-dependent and not good predictors for very short segments. CONTIN, which tends to provide the better estimate of turns, predicts about four times as many residues in this conformation as is predicted by the LINCOMB analyses. A contributing factor to this variability may be differences in the types of turns in the database as well as the number of residues considered in the definition of the turn.

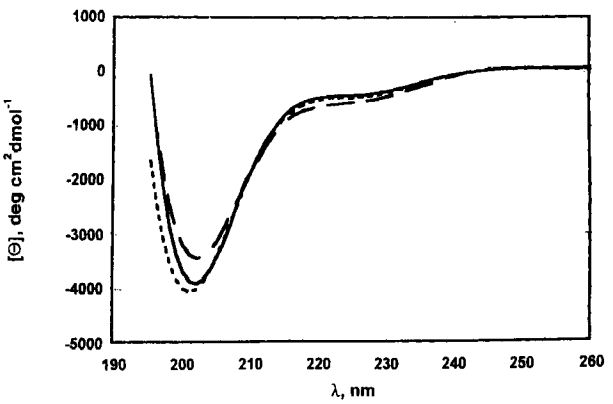

Fig. 8. Molar ellipticity [θ] as a function of wavelength for the far-UV CD of α_{s1}-casein (1–23) at 5°C, 10°C, and 25°C in 10 mM sodium phosphate–70 mM NaCl, pH 5.0. Concentration of peptide was 0.26 mg/ml. The mean residue ellipticities are expressed in deg · cm^2/dmole. —— 5°C - - - - 10°C — — — 25°C.

3.5. Effects of Temperature on CD and FTIR Spectra of α_{s1}-Casein (1–23)

Figure 8 shows the mean residue ellipticity as a function of wavelength in the far-UV for α_{s1}-casein (1–23) at 5°C, 10°C, and 25°C. The CONTIN analysis (Table V) of the CD spectra (Fig. 8) shows the absence of a significant temperature effect. In this temperature range, the conformation is about 32% β-turns, 30% β-sheet, and 13% α-helix or loop. The peptide appears to be stabilized by net positive charge and by turns, probably directed or promoted by proline residues, and contains a significant amount of predicted and persistent structure. Note that the trough at ~220 nm is somewhat similar to that observed for PPII and increases somewhat with temperature.

The secondary structure estimated from FTIR spectra in this temperature range is also given in Table VI. There is little change in the calculated amounts of turn, unordered, helix + loop, or sheet structure, indicating that local structures are stable through this temperature range. Both CD and FTIR data, therefore, suggest the presence of a significant amount of persistent structure.

3.6. Comparison of NMR Results for α_{s1}-Casein (1–23) with CD and FTIR Analyses

As noted above, the NMR data, especially the N–H chemical shift data, place this peptide in an unusual conformational niche. It is neither strongly hydrogen bonded to yield conventional helix and β-sheet nor a completely random coil. At 5°C, Table V shows good agreement between the CD and FTIR data for percentage of sheet, which would also be consistent with the N–H shifts (Fig. 6) that show a propensity for sheet (residues 6–10). Although NMR data did not predict turns, Fig. 6 indicates that Pro12 may interrupt a β-sheet like structure with a non-hydrogen-bonded turn. Assigning a turn to this segment of the peptide area is tenuous but consistent with the combined CD and FTIR data for the peptide as a whole, which predict 6–8 residues of turn. The data of Fig. 6 show the potential for an α-helical-like structure for residues 21–23. Such structures could be a portion of the loop structure suggested by FTIR and CD as well.

It was suggested by Tiffany and Krimm (1968) that charged peptides such as poly(L-lysine) and poly (L-glutamate) are not truly random coils but contain elements of a left-handed PPII helix (Siligardi and Drake, 1995). Two spectral features of poly-L-proline peptides are IR bands at 1650, 1636, and 1620 cm^{-1} (Dukor and Keiderling, 1991) and CD spectra somewhat similar to those shown in Fig. 8 for the casein peptide (Makarov *et al.*, 1984). The *ab initio* calculations of Torii and Tasumi (1998) suggest that such PPII-helical elements are left-handed and would have ϕ, ψ angles of $-100°$ and 170°, placing them in the upper left quadrant of a Ramachandran plot, near β-sheet. Thus, these elements would have N–H chemical shifts close to β-sheet and CD spectra very similar to those in Fig. 8. This would argue for a significant amount of polyamino acid-like structure for the peptide, i.e., a left-handed PPII helix (Siligardi and Drake, 1995). Such PPII structures are not usually H-bonded, as shown by the NOE data. A PPII type of structure could then be proposed for residues 1–6 containing Pro2 and Pro5 because PPII (n = 4) also has FTIR bands at 1620 cm^{-1}, as indicated in Table IV. Note that PPII ellipticities are not included in any of the CD basis sets, but the spectrum is similar to that of β-lipotropin (Makarov *et al.*, 1984). As noted above, Smyth *et al.* (2001) suggested the presence of PPII structures in casein by Raman optical activity studies.

3.7. Molecular Modeling

A significant decrease in energy resulted from solvating the blocked peptide containing residues 1–23 after it was excised from the refined model of α_{s1}-casein

Table VI. Effects of Temperature on CD and FTIR Secondary Estimates for α_{s1}-Casein (1–23)

Method	Temperature (°C)	β-Sheet (%)	Turns (%)	Unspecified (%)	α-Helix or loop (%)
CD[a]	5	29	31	26	13
	10	30	32	26	12
	25	30	32	25	13
FTIR[b]	5	26	39	11	23
	10	32	40	3	25
	25	25	39	11	30

[a] Average fits, six accumulations at 5°C, 10°C, and 25°C.
[b] 1400 accumulations at each temperature.

(Kumosinski *et al.*, 1994*c*), an approach that mimics the real peptide-water system investigated with NMR, CD, and FTIR. Note that energies for peptides refined in water are not comparable with those for the peptides refined *in vacuo* (Table VII). Subsequent molecular dynamics operations at increasing temperatures, each followed by a brief energy minimization, improved the energy of the peptide even more. The dynamics run at 250 K yielded the best total energy for the peptide–water system, −8203.1 kcal/mole, and the system reached equilibrium after 3000 psec. However, the total energy for the peptide alone was −1003.9 kcal/mole, indicating that solvation makes a very large contribution to the stability of the total system.

No major changes occurred in the overall backbone conformation of α_{s1}-casein (1–23) as a result of dynamics operations, as observed previously when the peptide was refined *in vacuo* at temperatures high enough to promote acquisition of a different peptide backbone conformation (Malin and Brown, 1995). Both *in vacuo* and in water, slight alterations in orientation of side chains or backbone conformations occurred. In contrast, simulations of several other α_{s1}-casein peptides, both *in vacuo* and solvated (Malin and Brown, 1995, 1999), resulted in significant conformation changes. These results suggest that the structure of α_{s1} (1–23) is inherently stable, cannot be induced to adopt a radically different conformation, and is essential to the stability of the casein framework in micelles, as discussed below.

The peptide contains three Pro residues, each of which is involved in a turn under SYBYL definitions. Figure 9 shows a contour Ramachandran plot (Morris *et al.*, 1992) for the refined model generated by Swiss-Pdb Viewer on the ExPASy Molecular Biology Server (http://www.expasy.ch/). The map shows Pro2 and Pro5 in one type of turn, whereas the turn of Pro12 is in a different class.

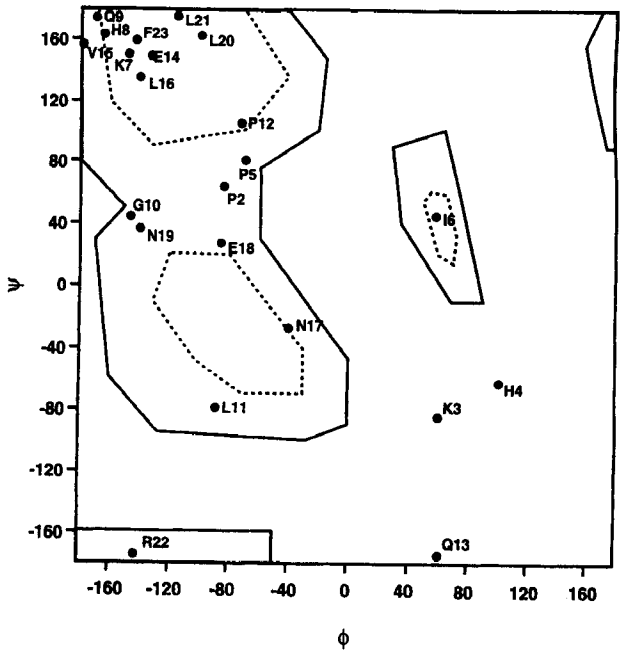

Fig. 9. Contour Ramachandran plot (Morris *et al.*, 1992), generated by Swiss-Pdb Viewer using ϕ, ψ angles for the solvated peptide, α_{s1}-casein (1–23), after refinement by molecular dynamics. Areas enclosed by dashed lines contain core residues, i.e., those most favored for specific types of defined secondary structure: β-sheet and some turns in the upper left quadrant, α-helix in the lower left quadrant, and other turns in the upper right quadrant. Areas enclosed by solid lines are considered allowed or within slightly less stringent definitions. The regions outside the solid lines are disallowed.

Table VII. Effect of Molecular Dynamics Refinement on Energies and Structural Features of α_{s1}-Casein (1–23)

	Solvated		*In vacuo*[a]	
	Initial	Final	*In vacuo*	With H_2O[b]
Energy (kcal/mole/residue)	15.34	−13.2	−55.2	−221.7
H bonds	—[c]	—[c]	15	28
C_α–C_α distance[d](Å)	18.3	17.0	18.2	15.5

[a] Malin and Brown (1995).
[b] *In vacuo* structure placed in a water droplet without further refinement.
[c] Hydrogen bonds cannot be calculated for solvated molecules.
[d] Distance between α-carbons of amino and carboxyl ends.

The bend promoted by Pro2 is close to the N-terminal and was altered somewhat during dynamics and minimizations (Fig. 10). Pro5 is preceded by Lys3 and His4, both of which are defined by SYBYL as participating in turns (Fig. 9). Based on the data of Table IV, another possible conformation for residues 1–6 would be a short PPII-like segment. The 1620-cm^{-1} band in the FTIR analysis and CD curves (Fig. 8) suggests the possibility of such a structure. In addition, optical Raman spectroscopy has indicated the presence of a significant amount of PPII in crude α_s-casein (Smyth *et al.*, 2001).

Lys7 is found in the β-sheet area of Fig. 9, with His8 nearby. These two residues appear to begin a short length of β-sheet-like structure that ends with Gly10. Gln13, located in a forbidden region of Fig. 9, appears at the beginning of a second length of predicted β-sheet structure, ending with Leu16. The latter and its neighbor, Asn17, also appear in a shallow bend, followed by Glu18 and Asn19 in an unordered conformation. Residues 20,

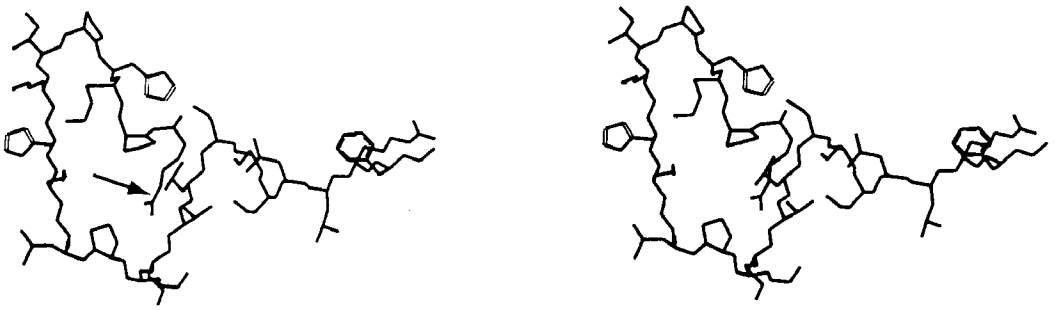

Fig. 10. Stereo representation of the refined, solvated peptide model, α_{s1}-casein (1–23). The carboxyl-terminal residue, Phe23, is at the right. The arrow points to the amino-terminal residue, Arg1.

21, and 23 of the refined peptide have ϕ, ψ angles located in the β-sheet area of a Ramachandran plot (Fig. 9).

The overall conformation of the simulated peptide, as analyzed by the SYBYL program, is in good agreement with the secondary structure predictions of the CD and FTIR analyses, as well as with NMR predictions. Six residues of the chain, 27%, have a β-sheet-like conformation, seven residues, 32%, are located in turns, and the remainder, 41%, are unordered (Table V).

4. CONCLUSIONS

Residues 14–23 of the α_{s1}-casein (1–23) peptide are a large part of the difference in sequence between α_{s1}-casein-B and the genetic variant, α_{s1}-casein-A, which lacks residues 14–26. This sequence difference dramatically alters the salt solubility and other physical properties of the protein. The Phe23–Phe24 bond cleaved by chymosin is therefore missing in α_{s1}-A. Plowman and Creamer (1995) showed that a β-extended structure ±4 residues from the cleaved bond is necessary to fit the chymosin active site, and the conformation of residues 20, 21, and 23 predicted by the model fits this constraint. After chymosin cleavage, these residues might adopt a nascent helix in solution, as suggested by N–H chemical shift data (Fig. 6) and possibly by CD and FTIR. The remainder of the peptide could be stabilized in solution by PPII-like structure and some extended sheet (residues 1–6 and 7–10, respectively). Pro12 and associated residues could represent a non-H-bonded turn.

Two important outcomes of this work are that stable secondary structures, detected by CD and FTIR and suggested by molecular modeling, occur in the peptide and by inference in the protein and that increased self-association of the peptide does not appreciably alter secondary structure. The bulk of the secondary structure defined by CD and FTIR for α_{s1}-casein (1–23) is present prior to, during, and after self-association. Therefore, we can rule out the possibility that secondary structure for this portion of α_{s1}-casein forms exclusively in response to formation of larger casein aggregates, as proposed by the rheomorphic hypothesis of Holt and Sawyer (1993).

The α_{s1}-casein (1–23) peptide is clearly not a random coil sampling all possible conformational space. It appears to maintain a relatively stable conformation in solution and may well exhibit the same stability as the N-terminal segment of the whole protein. The latter could be PPII-like, and this fact may account for its importance in casein–casein interactions; PPII regions are known to participate in protein–protein interactions such as signal transduction (Kay *et al.*, 2000). Moreover, conformationally stable peptides are generally characteristic of allergens, and this segment of α_{s1}-casein is known to be allergenic (Spuergin *et al.*, 1996). Finally, Horne (1998) has concluded that this region of α_{s1}-casein is important in its role in food systems as a surface-active agent. All of these selective roles for the N-terminal peptide argue for a persistent secondary structure.

ACKNOWLEDGMENTS

The authors thank Harold J. Dower for amino acid analysis and peptide sequencing, Dr. Robert L. Dudley for ^{19}F NMR analysis, Alicia Elsetinow for the CD spectra, and Joseph Unruh for the FTIR spectra.

REFERENCES

Adzhubei, A. A. and Sternberg, M. J. E. (1993). *J. Mol. Biol.* **229**, 472–493.

Alaimo, M. H., Wickham, E. D., and Farrell, H. M., Jr. (1999*a*). *Biochim. Biophys. Acta* **1431**, 395–409.

Alaimo, M. H., Farrell, H. M., Jr., and Germann, M. W. (1999*b*). *Biochim. Biophys. Acta* **1431**, 410–420.

Brahms, S. and Brahms, J. (1980). *J. Mol. Biol.* **138**, 149–178.

Byler, D. M., Farrell, H. M., Jr., and Susi, H. (1988). *J. Dairy Sci.* **71**, 2622–2629.

Chaplin, L. C., Clark, D. C., and Smith, L. J. (1988). *Biochim. Biophys. Acta* **1988**, 162–172.

Clark, M., Cramer, R. D., III, and Van Opdenbosch, N. (1989). *J. Computational Chem.* **10**, 982–1012.
Clarke, R. and Nakai, S. (1971). *Biochemistry* **10**, 3353–3357.
Creamer, L. K., Richardson, T., and Parry, D. A. D. (1981). *Arch. Biochem. Biophys.* **211**, 689–696.
Curley, D. M., Kumosinski, T. F., Unruh, J. J., and Farrell, H. M., Jr. (1998). *J. Dairy Sci.* **81**, 3154–3162.
Dukor, R. K. and Keiderling, T. A. (1991). *Biopolymers* **31**, 1747–1761.
Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., and Wright, P. E. (1988). *J. Mol. Biol.* **201**, 161–200.
Farrell, H. M., Jr., Kumosinski, T. F., Paulaski, P., and Thompson, M. P. (1988). *Arch. Biochem. Biophys.* **265**, 146–158.
Farrell, H. M., Jr., Malin, E. L., Brown, E. M., and Hoagland, P. D. (2001*a*). In *Calcium Binding Proteins* (Vogel, H., ed.), Humana Press, Totowa, NJ, Vol. I, Chapter 6, pp. 97–140.
Farrell, H. M., Jr., Wickham, E. D., Unruh, J. J., Qi, P. X., and Hoagland, P. D. (2001*b*). *Food Hydrocolloids,* in press.
Feeney, J., Roberts, G. C. K., Brown, J. P., Burgen, A. S. V., and Gregory, H. (1972). *J. Chem. Soc. Perkin Trans. II* **1972**, 601–604.
Fox, P. F., Singh, T. K., and McSweeney, P. L. H. (1995). In *Chemistry of Structure–Function Relationships in Cheese* (Malin, E. L., and Tunick, M. H., eds.), Plenum Press, New York, pp. 59–98.
Greenfield, N. J. (1996). *Anal. Biochem.* **235**, 1–10.
Holt, C. (1992). *Adv. Protein Chem.* **43**, 63–113.
Holt, C. and Sawyer, L. (1993). *J. Chem. Soc. Faraday Trans.* **89**, 2683–2690.
Horne, D. S. (1998). *Int. Dairy J.* **8**, 171–177.
Huq, N. L., Cross, K. J., and Reynolds, E. C. (1995). *Biochim. Biophys. Acta* **1247**, 201–208.
Kakalis, L. T., Kumosinski, T. F., and Farrell, H. M., Jr. (1990). *Biophys. Chem.* **38**, 87–98.
Kay, B. K., Williamson, M. P., and Sudol, M. (2000). *FASEB J.* **14**, 231–241.
Krimm, S. and Bandekar, J. (1986). *Adv. Protein Chem.* **38**, 181–481.
Kumosinski, T. F., and Farrell, H. M., Jr. (1991). *J. Protein Chem.* **10**, 3–16.
Kumosinski, T. F. and Unruh, J. J. (1996). *Talanta* **43**, 199–219.
Kumosinski, T. F., Brown, E. M., and Farrell, H. M., Jr. (1993*a*). *J. Dairy Sci.* **76**, 931–945.
Kumosinski, T. F., Brown, E. M., and Farrell, H. M., Jr. (1993*b*). *J. Dairy Sci.* **76**, 2507–2520.
Kumosinski, T. F., King, G., and Farrell, H. M., Jr. (1994*a*). *J. Protein Chem.* **13**, 681–700.
Kumosinski, T. F., King, G., and Farrell, H. M., Jr. (1994*b*). *J. Protein Chem.* **13**, 701–714.
Kumosinski, T. F., Brown, E. M., and Farrell, H. M., Jr. (1994*c*). In *Molecular Modeling from Virtual Tools to Real Problems* (Kumosinski, T. F., and Liebman, M. N., eds.), American Chemical Society, Washington, D. C., pp. 368–389.
Kumosinski, T. F., Unruh, J. J., and Farrell, H. M., Jr. (1997). *Talanta* **44**, 1441–1445.
Kumosinski, T. F., Uknalis, J., Cooke, P. H., and Farrell, H. M., Jr. (1996). *Lebensm. Wiss. Technol.* **29**, 326–333.
Makarov, A. A., Esipova, N. G., Lobachov, V. M., Girshovsky, B. A., and Pankov, Y. A. (1984). *Biopolymers* **23**, 5–22.
Malin, E. L. and Brown, E. M. (1995). In *Chemistry of Structure–Function Relationships in Cheese* (Malin, E. L., and Tunick, M. H., eds.), Plenum Press, New York, pp. 303–310.
Malin, E. L. and Brown, E. M. (1999). *Int. Dairy J.* **9**, 207–213.
Manavalan, P. and Johnson, W. C., Jr. (1987). *Anal. Biochem.* **167**, 76–85.
Marion, D. and Wüthrich, K. (1983). *Biochim. Biophys. Res. Commun.* **113**, 967–974.
Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992). *Proteins Struct. Func. Genet.* **12**, 345–364.
Pedersen, J. T. and Moult, J. (1995). *Proteins* **23**, 454–460.
Pederson, J. A., Mileski, G. J., Weimer, B. C., and Steele, J. L. (1999). *J. Bacteriol.* **181**, 4592–4597.
Plowman, J. E. and Creamer, L. K. (1995). *J. Dairy Res.* **62**, 451–467.
Provencher, S. W. and Glöckner, J. (1981). *Biochemistry* **20**, 33–37.
Rollema, H. S. and Brinkhuis, J. A. (1989). *J. Dairy Res.* **56**, 417–425.
Schachman, H. K. (1959). *Ultracentrifugation in Biochemistry,* Academic Press, New York.
Schmidt, D. G. (1982). In *Developments in Dairy Chemistry* (Fox, P. F., ed.), Applied Science, Essex, England, Vol. 1, pp. 61–86.
Siligardi, G. and Drake, A. F. (1995). *Biopolymers* **37**, 281–292.
Smyth, E., Syme, C. D., Blanch, E. W., Hecht, L., Vasak, M., and Barron, L. D. (2001). *Biopolymers* **58**, 138–151.
Spuergin, P., Mueller, H., Walter, M., Schiltz, E., and Forster, J. (1996). *Allergy* **51**, 306–312.
Tanford, C. (1967). *Physical Chemistry of Macromolecules,* Wiley, New York.
Tiffany, M. L. and Krimm, S. 1968. *Biopolymers* **6**, 1379–1382.
Torii, H. and Tasumi, M. (1998). *J. Raman Spectrosc.* **29**, 81–86.
Tsuda, S., Niki, R., Kuwata, T., Tanaka, I., and Hikichi, K. (1991). *Magnetic Resonance Chem.* **28**, 1097–1102.
Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Algona, G., Profeta, S., Jr., and Weiner, P. (1984). *J. Am. Chem. Soc.* **106**, 765–784.
Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986). *J. Computational Chem.* **2**, 230–252.
Wider, G. and Wüthrich, K. (1993). *J. Magnetic Resonance* **102B**, 239–241.
Wilder, C. L., Friedrich, A. D., Potts, R. O., Daumy, G. O., and Francocur, M. L. (1992). *Biochemistry* **31**, 27–31.
Wishart, D. S. and Sykes, B. D. (1994). *Meth. Enzymol.* **239**, 363–392.
Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991). *J. Mol. Biol.* **222**, 311–333.
Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992). *Biochemistry* **31**, 1647–1651.
Yang, J. T., Wu, C. S., and Martinez. H. M. (1986). *Meth. Enzymol.* **130**, 208–269.